An introduction to **individual-based models** in epidemiology, and to inform health policy

Loveleen Bansi-Matharu
Valentina Cambiano
Paul Revill*
Andrew Phillips (presenter)

UCL, London, UK and *University of York, York, UK

Health policies relate to how we organize our health care system or promote healthy life choices in order to generate health in our populations.

Let’s imagine our aim is to maximise overall quality and length of life in our population.

Imagine we are considering a change in health policy.

What would we want to know ?

Whether the policy is likely to lead to an increase in overall quality and length of life in our population.

How do we go about finding out if the policy is likely to lead to an increase in overall quality and length of life in our population ?

Often a new policy will be to introduce a new “intervention”, such as a new drug to prevent or treat a disease, or a new test for diagnosing or predicting risk of a disease earlier.

We need to consider the direct and indirect effects of the policy on the population.

Often the health effects of a proposed intervention are entirely positive.

However, most health care systems have limited resources. Introduction of the intervention may well result in less resources for other existing health care.

The policy decision is about whether the result of the introduction of the intervention would be a net gain in health.

In addition to these resource-use considerations, for some potential new interventions, the health benefits of the intervention may be potentially offset by negative health effects.

For example, there can be adverse effects of drugs.

In the case of drugs for infectious diseases one such potential adverse effect is the fact that the policy to use the drug in a given way may result in increases in spread of drug resistant infections.

So how do we go about addressing these questions so that we can help to inform health policy ?

Often the persons responsible for making policy may have to use their best judgement after consulting with various experts.

A disadvantage of this approach is that the decision process is not transparent.

Often the decision can be far from straightforward and involve trade offs between positive health effects, direct negative effects, and indirect negative effects due to resources being diverted.

It is difficult for a policy-maker to appropriately balance all these considerations.

Another approach is to try to represent the comparison of the policy options mathematically. This might be referred to as an “analytic framework”.

Various approaches are used. e.g. Decision analysis models / decision trees, Markov models, compartmental models, **individual-based** models

We focus in the rest of this presentation on **individual-based** models

An **individual-based** model is designed to simulate experiences of people in a population over time

Each time the model is run it generates a simulated “data set” of variable values for a cohort of individuals representing the population of interest

Example

Status of a population at a certain time

Person	Var 1	Var 2	Var 3	Alive / dead
1				
2				
3				
4				
5				
.				
.				
9999				
10,000				

Example

Status of a population at a certain time

Person	Var 1	Var 2	Var 3	Alive / dead
1				
2				
3				
4				
5				
.				
.				
9999				
10,000				

Status of the population 3 months later

(Note that any time step length can be chosen)

Person	Var 1	Var 2	Var 3	Alive / dead
1				
2				
3				
4				
5				
.				
.				
9999				
10,000				

Example: simple model of cardiovascular disease

Status of a population at a certain time

Person	SBP (mm Hg)	Total cholesterol (mmol/l)	Smoker	Alive / dead
1	120	4.5	Yes	Alive
2	180	3.8	No	Alive
3	110	2.9	No	Alive
4	100	6.1	No	Alive
5	130	4.7	No	Alive
.	.	.	.	.
.	.	.	.	.
9999	145	2.8	No	Alive
10,000	105	6.4	Yes	Alive

Status of a population 3 months later

Person	SBP (mm Hg)	Total cholesterol (mmol/l)	Smoker	Alive / dead
1	120	4.3	Yes	Alive
2	185	3.9	No	Alive
3	110	2.9	No	Alive
4	105	6.1	No	Alive
5	125	4.8	No	Alive
.	.	.	.	.
.	.	.	.	.
9999	145	3.0	No	Alive
10,000	105	6.4	No	Alive

Q1 2010

Person	Var 1	Var 2	Var 3	Alive / dead
1				
2				
3				
4				
5				
.				
.				
9999				
10,000				

Q2 2010

Person	Var 1	Var 2	Var 3	Alive / dead
1				
2				
3				
4				
5				
.				
.				
9999				
10,000				

Q3 2010

Person	Var 1	Var 2	Var 3	Alive / dead
1				
2				
3				
4				
5				
.				
.				
9999				
10,000				

Q4 2010

Person	Var 1	Var 2	Var 3	Alive / dead
1				
2				
3				
4				
5				
.				
.				
9999				
10,000				

Q1 2011

Person	Var 1	Var 2	Var 3	Alive / dead
1				
2				
3				
4				
5				
.				
.				
9999				
10,000				

Q2 2011

Person	Var 1	Var 2	Var 3	Alive / dead
1				
2				
3				
4				
5				
.				
.				
9999				
10,000				

Q3 2011

Person	Var 1	Var 2	Var 3	Alive / dead
1				
2				
3				
4				
5				
.				
.				
9999				
10,000				

Q4 2011

Person	Var 1	Var 2	Var 3	Alive / dead
1				
2				
3				
4				
5				
.				
.				
9999				
10,000				

Q1 2012

Person	Var 1	Var 2	Var 3	Alive / dead
1				
2				
3				
4				
5				
.				
.				
9999				
10,000				

Q2 2012

Person	Var 1	Var 2	Var 3	Alive / dead
1				
2				
3				
4				
5				
.				
.				
9999				
10,000				

So, for example, if we want to represent values of 100 variable on 10,000 individuals as they change every 3 months from 2010 to 2024 this means when we run the model we will be generating up to

100 x 10,000 variable values per 3 months

x 60 3-month periods from 2010 – 2024

= 60 million variable values

(although note that we do not need to create a new variable value unless the value changes)

Note that any time step length can be chosen

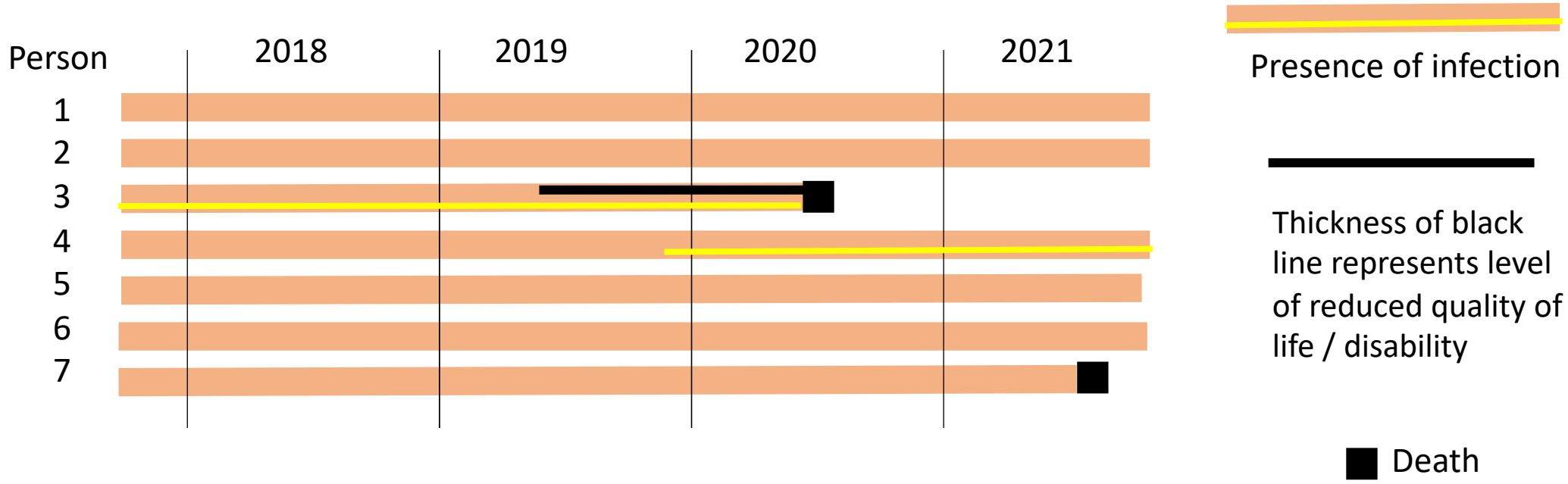
In some **individual-based** models, the experience of each individual is entirely independent of the other individuals.

In other models, a person's health experience can depend on others. A common example of this is in modelling of **infectious diseases**.

But there are also other examples, such as when a person's health choices (e.g. diet) are influenced by others in their social network.

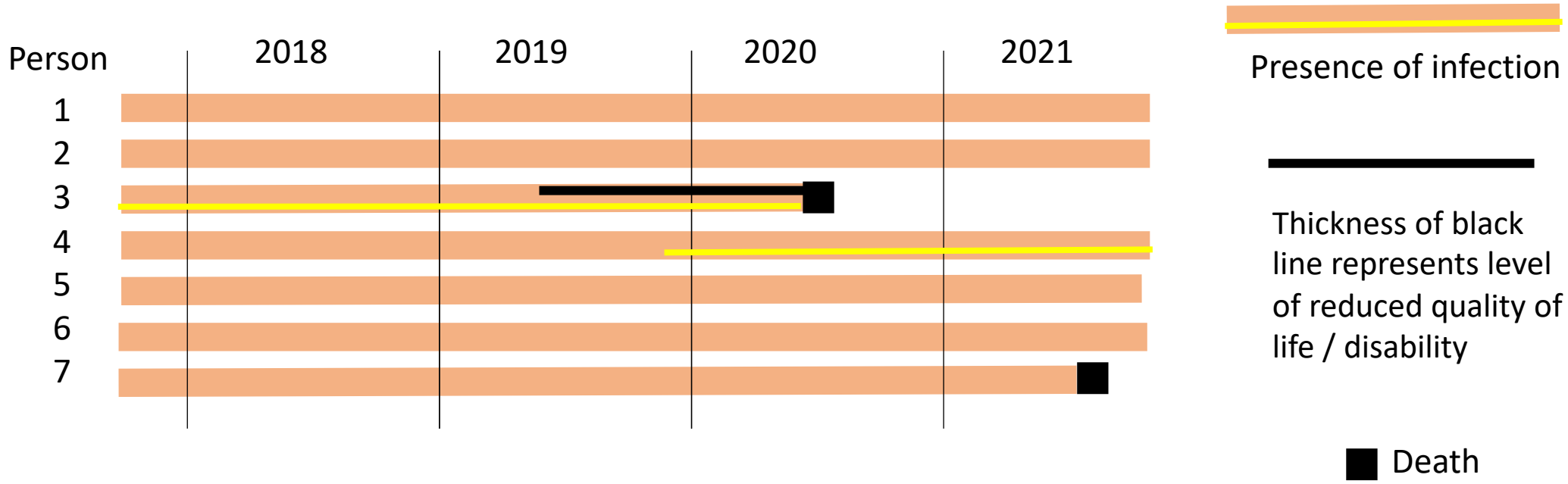
Modelling infection within a population

Here we represent the life years experienced by each individual with horizontal orange lines



We consider the outcomes of a whole population of people, including those who could potentially be susceptible to infection as well as those already infected.

Modelling infection within a population



We model how infection is transmitted and newly acquired within the population by modelling determinants of infectivity of infected people and susceptibility of uninfected people.

These determinants depend on the infection; e.g. for HIV, depends on condomless sex and viral load of infected partner

Example: Simple model of an infection

Follow this link https://colab.research.google.com/drive/1P9hdU_49VonGa_RycHhH0ienshtJnaVC to see an example of an individual based model of an infection. This is coded in Python.

Press the >  to run

You will see a series of tables tracking the experiences of a cohort of 1000 people over 100 days.

The columns show:

day	**person ID**	**whether infected** (and infectious)	**contacts** (others in the cohort with whom had contact)	**infectious contacts** (contacts who are infected / infectious)	**ever infected**	**day of infection** (if ever infected)

Initially only person 0 is infected.

Look at the people who had contact with person 0. Look at the next day for these people to see if they became infected (the probability is 0.2).

Example: Simple model of an infection

At the bottom of the program you will see that you can change the following to produce different epidemic patterns.

Number of days to run
Population size
Initial number of infected
Chance to become infected per infectious contact
Number of days to remain infected / infectious
Number of contacts / interactions each day is set to interaction_multiplier * population size

This is an example of an individual-based model in which there is an explicit network of contact patterns between individuals.

An alternative approach is where the risk of acquiring an infection for a subject through a given contact is determined at random based on the concurrent prevalence of infectious contacts amongst contacts to whom the subject is potentially exposed.

An individual based model is largely defined by:

- The list of variables / attributes for each individual

- The expressions that determine how each variable value is updated in each time step

An individual based model is largely defined by:

- The list of variables / attributes for each individual

- The expressions that determine how each variable value is updated in each time step

Note that the expressions which represent the development and progression of a disease are what we hypothesize to be true underlying causal effects.

Examples of variables / attributes for each individual

Age
Gender
Education
Respiratory infection contact behaviour
Enteric infection contact behaviour
Sexual infection contact behaviour
Exposure to environmental pathogens
Presence of infection with various infectious pathogens
Host genetic factors (not updated over time)
Nutritional intake
Dietary intake
Biomarkers, including biomarkers indicating pathology
Indicators of organ function
Presence of clinically manifest disease
Whether a given condition has been diagnosed
Treatments administered
DALYs incurred in this period

Expressions that determine how a variable value is updated in each time step: examples

Presence of HIV (in a model of sexual transmission):

Updating for a person without HIV:

If has no condomless sex in this time step: 0 risk of infection

If has condomless sex with a new partner (of HIV status unknown to the person) and with no other partner:

risk of infection in next period depends on HIV prevalence (and in those with HIV, the viral load distribution) amongst others in the population having condomless sex with new partners in the person's age range.

We might want this to also depend on the number of times in the time period that they had condomless sex.

Expressions that determine how a variable value is updated in each time step: examples

Systolic blood pressure:

Updated value might depend on:

Existing SBP value
Age
Gender
Body mass index
Exercise level
Salt intake
Whether taking an anti-hypertensive drug

It may be, for example, that these determine the probability of a 5mmHg increase and whether or not this increase occurs is determined by random chance.

Note that we are generating values for the true underlying systolic blood pressure, not the measured blood pressure

If we want to take into account the error in measurement of blood pressure we can create a separate variable which represents the measured value at any point in time

This will be the true value plus a randomly determined error

The same principle applies to **any diagnostic test for any condition** – we use our knowledge of the accuracy of the test to determine who is diagnosed, who is not diagnosed despite having the condition, and who is falsely diagnosed.

The accuracy of the test is specified in terms of its **sensitivity** and its **specificity.** Sensitivity is about how good the test is in picking up the disease in people who have it. The specificity is about how good the test is at avoiding mistakes and diagnosing people with the disease when they don't have it.

Expressions that determine how a variable value is updated in each time step: examples

Occurrence of a myocardial infarction:

Might depend on:

Age
Gender
Current systolic blood pressure
Smoking status
LDL cholesterol level
Presence of diabetes
Pre-existing cardiovascular disease
+

These variables might determine the risk of an event and whether the event actually occurs depends on a random draw of a number between 0 and 1 where there is equal probability of each value (which is called a Uniform distribution).

Expressions that determine how a variable value is updated in each time step: examples

Occurrence of a myocardial infarction:

So we might have an expression

Risk of myocardial infarction in the next time step =

p1 + p2 x age + p3 if male + p4 x SBP + p5 if a current smoker + p6 x LDL cholesterol + p7 if diabetes + p8 if pre-existing CVD + a random number drawn at random from a specified distribution

p1 to p8 are examples of model **parameters**

(We are in fact more likely to have an expression on the logarithmic scale
and then transform to the linear scale)

Model calibration (1)

“Calibration” is the process of trying to ensure that the model reflects reality, in so far as a model can be expected to do.

This involves comparing model outputs with data.

Since the model simulates changes in variable values for individuals over time it is possible to compare model outputs with any source of data by mimicking the way the study collected data.

For example, a cross sectional study can be mimicked in the simulated outputs and results compared with the actual study. Similarly for cohort studies.

Model calibration (2)

Calibration might consist of running the model multiple times and each time sampling the value for each parameter from a distribution reflecting our uncertainty over that parameter value and possible variability between settings in the value.

Then we might compare the model outputs with data. We might exclude those model runs for which the calibration is judged to be inadequate because they do not appear to plausibly reflect reality in our settings of interest.

The sets of parameter values that result in model outputs that look close to the data might be preserved for future runs of the model when evaluating policies.

Dealing with uncertainty

Usually, there will be **uncertainty or variability** between settings in appropriate parameter values used in expressions to update variable values over time.

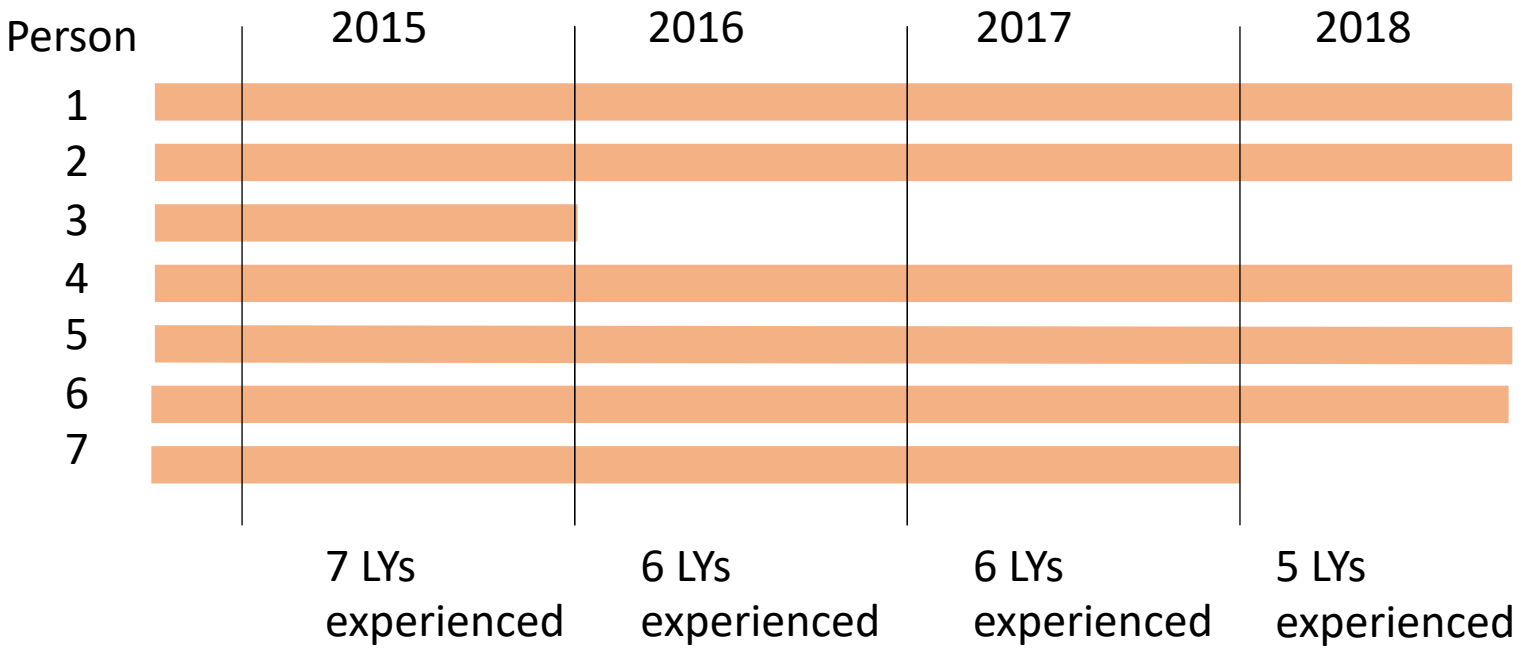
In this case it is again helpful to choose a distribution of values for each parameter rather than single fixed value. A value for each parameter is selected each time the model is run.

It may be that some model runs produce outputs that are outside our acceptable ranges calibration, and these can be discarded / ignored.

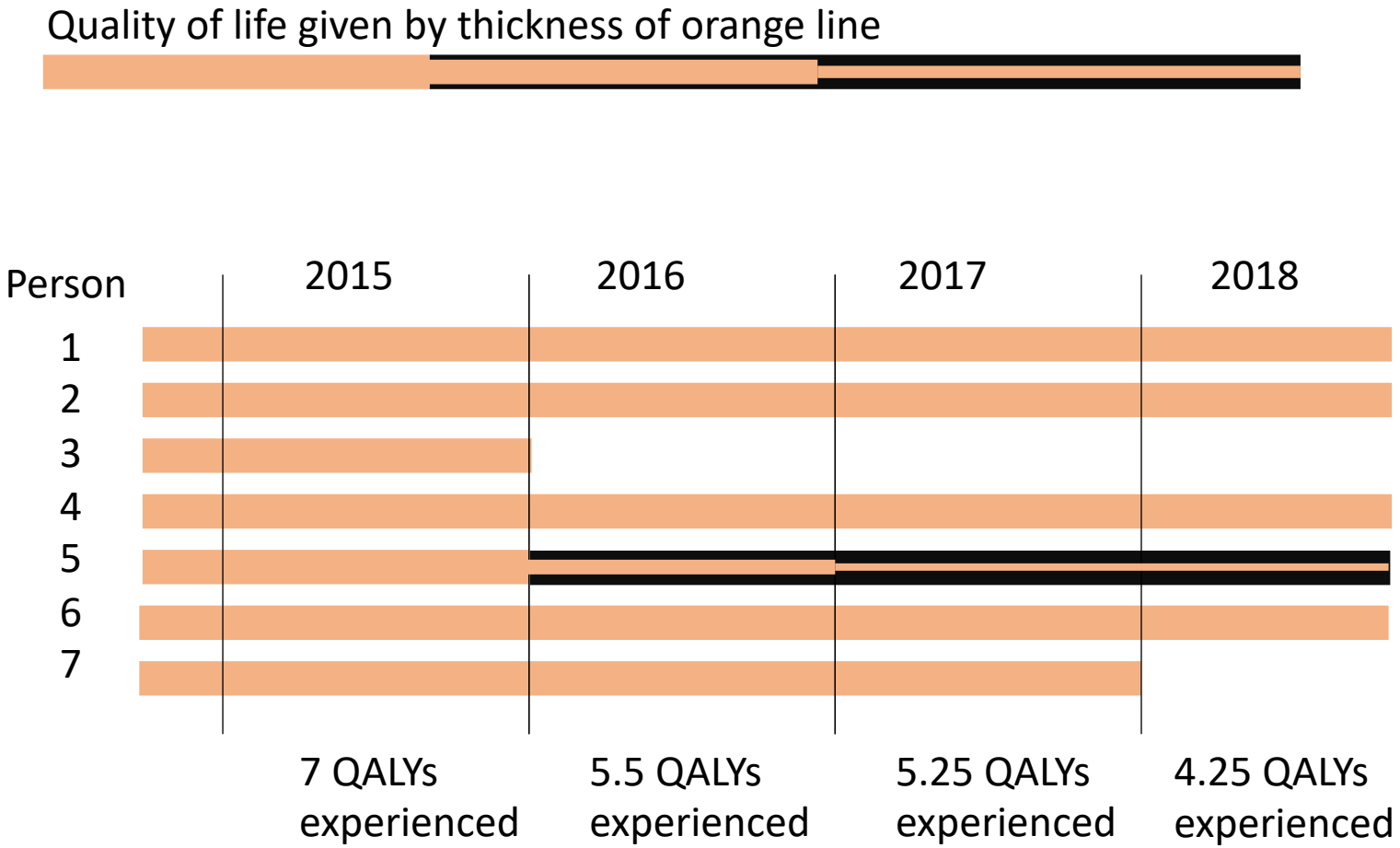
By running the simulations of the model multiple times at looking at the answer to the policy question each time, this can allow conveying of uncertainty in the answer to the policy question. It can also allow analyses to study how each parameter influences the answer to the policy question.

When calculating health outcomes resulting from implementation of a policy, how do we measure health ?

Life years


Person
2015
2016
2017
2018
1
2
3
4
5
6
7
7 LYs experienced
6 LYs experienced
6 LYs experienced
5 LYs experienced

QALYs – Quality adjusted life years

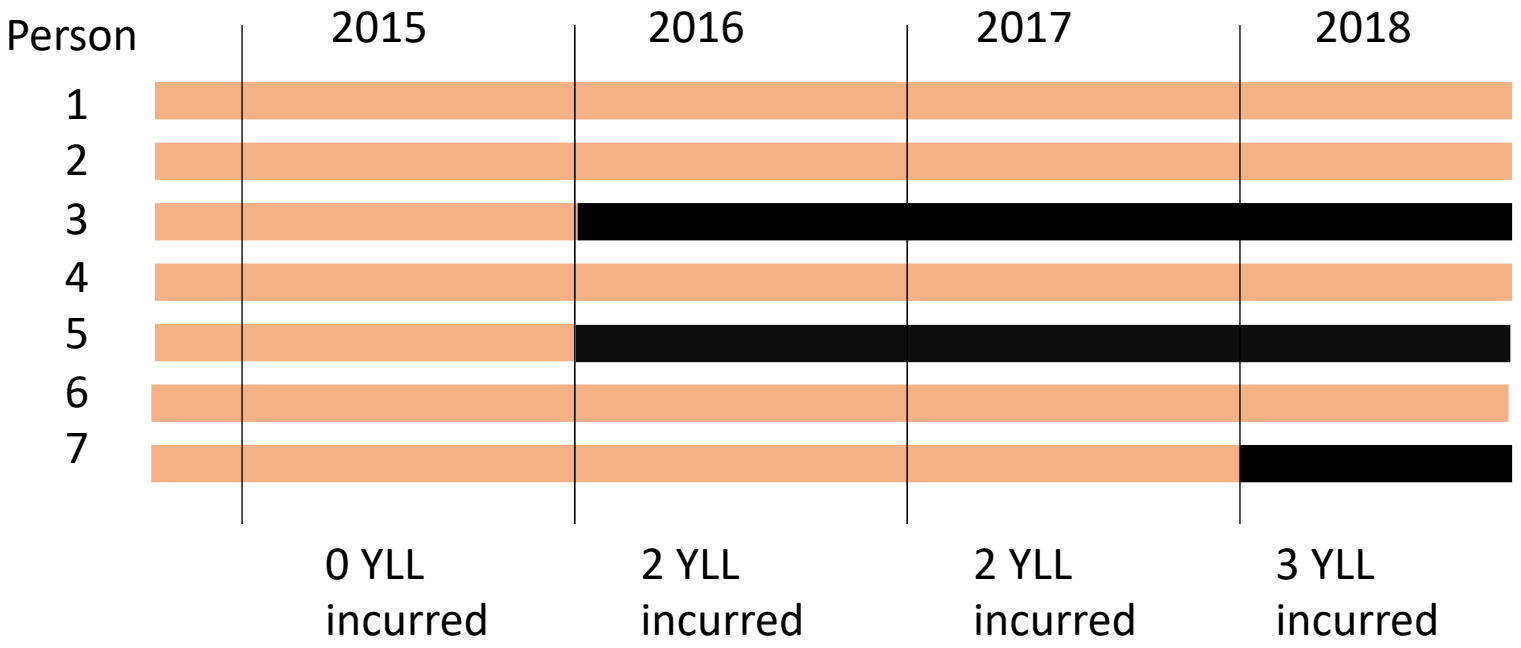


(QALYs incurred given by width of orange line)

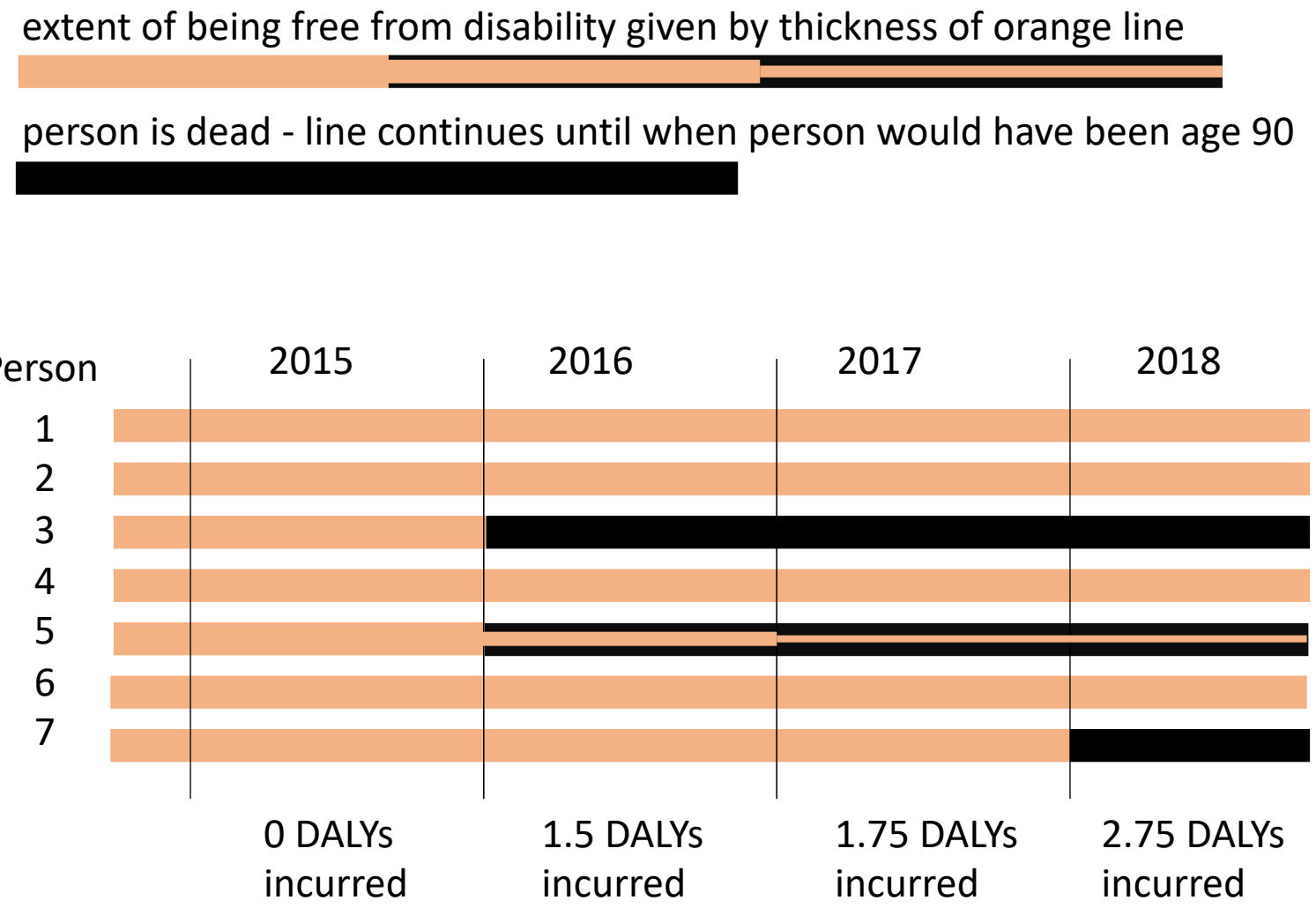
Years of lost life (YLL) (before age 90)

person is alive

person is dead - line continues until when person would have been age 90



DALYs - Disability adjusted life years



(DALYs incurred given by width of black line)

How to compare the effects of two health policies (1)

- Run the model up to the time point at which the choice of policy arises
- Then run the model from this policy choice time point into the future, separately for each of the two health policies
- Compare outcomes, including DALYs and costs
- Can then calculate the difference in costs and DALYs

How to compare the effects of two health policies

- If the new policy results in lower DALYs but also higher costs:

$$\text{Incremental cost-effectiveness ratio (ICER)} = \frac{\text{difference in costs}}{\text{DALYs averted}}$$

- This ICER can be compared against a pre-defined cost-effectiveness threshold which represents the added cost to the health care system to avert one further DALY. If the ICER is below the threshold this suggests the new policy is cost-effective.

Individual-based models also have uses in descriptive epidemiology

Example: life expectancy of a person infected with HIV at age 30

How to go about estimating this ?

Life expectancy is often calculated by assuming that age-specific death rates at a given point in time hold for a person's lifetime

But in a person with HIV the risk of death can change dramatically over time.

In a person who is not on treatment (e.g. due to not being diagnosed) the rate of death increases sharply as the CD4 count declines to very low levels. The rate of decline in CD4 count varies a lot between individuals.

Example: life expectancy of a person infected with HIV at age 30 …. continued

If the person is tested for HIV and diagnosed and starts treatment their risk of death declines substantially.

In a person on treatment with an undetectable HIV viral load and a high CD4 count, the extent to which risk of death is higher than if HIV had not been present is low.

So let's consider how we might estimate life expectancy using an individual-based model.

Let's consider a person who is infected with HIV at age 30

Recall that an individual based model is largely defined by:

- The list of variables / attributes for each individual
- The expressions that determine how each variable value is updated in each time step

Example: life expectancy of a person infected with HIV at age 30 …. continued

- **The list of variables / attributes for each individual**

  For example…..

  Viral load
  CD4 count
  diagnosed (yes/no)
  on antiretroviral treatment (yes / no)
  dead from AIDS
  dead from non-AIDS

Example: life expectancy of a person infected with HIV at age 30 …. continued

- **The expressions that determine how each variable value is updated in each time step**

  - Viral load at time t+1 depends on
    - viral load at time t
    - whether on treatment
    - random chance
  - CD4 count at time t+1 depends on
    - viral load at time t
    - CD4 count at time t
    - whether on treatment
    - random chance
  - Diagnosed at time t+1 depends on
    - whether already diagnosed at t
    - whether diagnosis occurs at t

Example: life expectancy of a person infected with HIV at age 30 …. continued

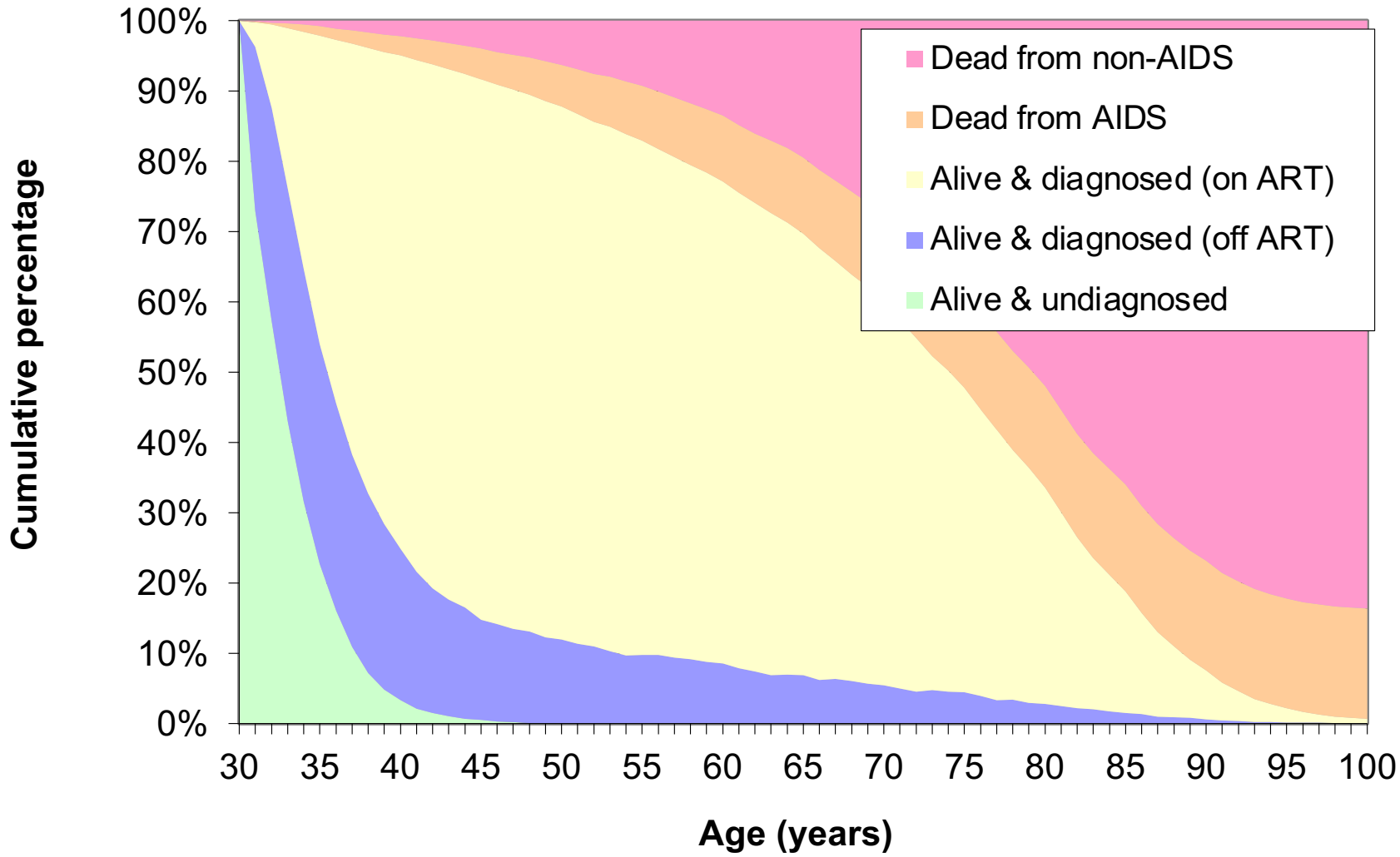
- **The expressions that determine how each variable value is updated in each time step**

- Whether on treatment at time t+1 depends on	whether on treatment at time t rate of treatment interruption rate of treatment resumption random chance
- Whether dead from AIDS at time t+1 depends on	viral load at time t CD4 count at time t whether on treatment random chance
- Whether dead from non-AIDS at t+1 depends on	age random chance

Example: life expectancy of a person infected with HIV at age 30 …. continued

- By building an individual-based model with these variables and transition rules and running it for 70 years for 1000 simulated people we get a picture of the various possible outcomes for the person by different ages……

Example: life expectancy of a person infected with HIV at age 30



Nakagawa et al, AIDS, 2012

https://journals.lww.com/aidsonline/Fulltext/2012/01280/Projected_life_expectancy_of_people_with_HIV.9.aspx

Example: life expectancy of a person infected with HIV at age 30 …. continued

- This is more complex in reality because of a need to account for specific drugs and drug resistance and there is likely a slightly increased risk of non-AIDS death in people with HIV. (This was taken account of in the cited paper.) But this illustrates the approach.

How are individual-based models built and run ?

These are often programmed in a basic language such as C++ or Python but can also be done using a statistical package such as SAS

There is also dedicated software.

For example, Anylogic
https://www.anylogic.com/use-of-simulation/agent-based-modeling/

June
https://royalsocietypublishing.org/doi/10.1098/rsos.210506

https://www.theoj.org/joss-papers/joss.03539/10.21105.joss.03539.pdf

Examples of individual based models

HIV

http://hivmodeling.org/model-database/emod-hiv
http://hivmodeling.org/model-database/hiv-synthesis
http://hivmodeling.org/model-database/popart-ibm
http://hivmodeling.org/model-database/thanzi-la-onse-tlo

COVID-19

https://github.com/institutefordiseasemodeling/covasim.
https://github.com/BDI-pathogens/OpenABM-Covid19

Further reading

Willem, L., Verelst, F., Bilcke, J. *et al.* Lessons from a decade of individual-based models for infectious disease transmission: a systematic review (2006-2015). *BMC Infect Dis* **17,** 612 (2017). https://doi.org/10.1186/s12879-017-2699-8

Roche B, Duboz R. Individual-Based Models for Public Health. Handbook of Statistics. 2017;37:347–65. doi: 10.1016/bs.host.2017.08.008. Epub 2017 Oct 10. PMCID: PMC7148902.

Summing Up

We aimed to provide an introduction to individual-based modelling in epidemiology and to inform policy.

Please post any comments with suggestions for enhancements.